Realistic medical image super-resolution with pyramidal feature multi-distillation networks for intelligent healthcare systems

Sheng Ren[1,2] · Kehua Guo[1] · Jianguang Ma[3] · Feihong Zhu[1] · Bin Hu[1] · Haoming Zhou[4]

Abstract

There are two key requirements for medical lesion image super-resolution reconstruction in intelligent healthcare systems: clarity and reality. Because only clear and real super-resolution medical images can effectively help doctors observe the lesions of the disease. The existing super-resolution methods based on pixel space optimization often lack high-frequency details which result in blurred detail features and unclear visual perception. Also, the super-resolution methods based on feature space optimization usually have artifacts or structural deformation in the generated image. This paper proposes a novel pyramidal feature multi-distillation network for super-resolution reconstruction of medical images in intelligent healthcare systems. Firstly, we design a multi-distillation block that combines pyramidal convolution and shallow residual block. Secondly, we construct a two-branch super-resolution network to optimize the visual perception quality of the super-resolution branch by fusing the information of the gradient map branch. Finally, we combine contextual loss and L1 loss in the gradient map branch to optimize the quality of visual perception and design the information entropy contrast-aware channel attention to give different weights to the feature map. Besides, we use an arbitrary scale upsampler to achieve super-resolution reconstruction at any scale factor. The experimental results show that the proposed super-resolution reconstruction method achieves superior performance compared to other methods in this work.

Keywords Medical image · Multi-distillation · Pyramidal feature · Super-resolution

1 Introduction

With the development of deep learning and the internet of things (IoT), the increase in visualization data, and the improvement of computing power, medical science and technology are coming together to provide better healthcare services. In IoT-based intelligent healthcare systems, more than 90% of medical data are medical images, mainly including X-ray imaging (X-CT), magnetic resonance imaging (MRI), nuclear medicine imaging (NMI), and ultrasound imaging (UI). Medical imaging data have become an important basis for doctors to diagnose diseases. The use of deep learning technology to analyze and process medical images can assist doctors in qualitative and even quantitative analysis of lesions in medical image reconstruction, automatic marking, recognition, and annotation, thereby effectively improving the accuracy and reliability of medical diagnosis [1, 2]. It is of great significance to alleviate the scarcity and uneven distribution of medical resources. Intelligent healthcare systems based on deep learning and IoT have been applied in related fields [3, 4]. For example, Philips' OncoSuite, an application program for intelligent tumor interventional therapy, can optimize tumor focus display, guide catheter placement, treatment, and efficacy evaluation, and other treatment links, making tumor interventional therapy more rational and standardized. Huawei's artificial intelligence medical image

✉ Kehua Guo
guokehua@csu.edu.cn

✉ Jianguang Ma
jianguangma1970@126.com

1 School of Computer Science and Engineering, Central South University, Changsha 410083, China

2 School of Computer and Electrical Engineering, Hunan University of Arts and Science, Changde 415000, China

3 Institute of Systems Engineering, Academy of Military Sciences, People's Liberation Army, Beijing 100000, China

4 Donald Bren School of Information and Computer Sciences, University of California, Irvine, CA 92604, USA

analysis technology can assist doctors in cervical cancer screening, stroke segmentation, and automatic generation of plain film diagnostic reports, improving the efficiency and accuracy of disease diagnosis. In 2019, Nature Medicine published research from Guangzhou Medical University and the University of California on the implementation status and future development of AI technology in the medical and health field. It is believed that medical imaging-based radiology, pathology, ophthalmology, and dermatology combining computer vision technology to achieve better automatic analysis or diagnosis prediction will be the first clinical field to realize the transformation of AI technology [5, 6]. In response to COVID-19, Huawei has developed a telemedicine platform, which can not only improve the efficiency and effectiveness of diagnosis, but also effectively reduce the risk of infection through remote online consultations with experts. The platform was deployed in Thailand in March 2020 to help Thai people fight COVID-19. Intelligent healthcare systems based on deep learning and IoT require high-resolution (HR) medical images to assist diagnosis. However, interference from equipment technology, hardware cost, network bandwidth, shooting environment, and human factors, some medical images have low-resolution (LR). LR medical images lack high-frequency detailed information, and it is difficult to identify lesions, which is not conducive to assisting doctors in diagnosing diseases. Therefore, we study the super-resolution (SR) method based on medical images to restore LR medical images to HR images with rich high-frequency details and clear visual perception.

SR technology reconstructs the corresponding HR image based on the LR image, which is widely used in medical diagnosis, satellite communications, and other fields [7]. Traditional SR methods are mostly based on interpolation or example-based methods. The SR images reconstructed by these methods have a serious loss of edge and texture information and blurred visual perception. The Chinese University of Hong Kong used convolutional neural networks in SR for the first time and reconstructed SR images by constructing a three-layer neural network model [8–10]. SR convolutional neural networks (SRCNN) surpass traditional SR methods in reconstruction quality and efficiency. Enhanced deep super-resolution network (EDSR) introduces a residual structure to enhance the depth of the SR network level, which effectively improves the quality of SR reconstruction. Residual feature aggregation network for image super-resolution (RFASR) uses the hierarchical features on the residual branch to enhance the quality of SR reconstruction. SRCNN, EDSR, and RFASR are SR methods based on pixel space optimization, and the reconstructed SR image often lacks high-frequency information, resulting in unclear visual perception. To improve the quality of visual perception, Twitter has applied generative adversarial networks (GAN) to solve the super-resolution problem [11], and proposed the super-resolution using a generative adversarial network (SRGAN). SRGAN uses content and adversarial loss to improve the realism of the restored picture. Enhanced SRGAN (ESRGAN) further improves the SR image quality by introducing the relativistic average discriminator, improving the network structure, and optimizing the loss function, and obtaining a better visual quality on the more realistic and natural texture [12]. However, these SR methods lack a comprehensive consideration of visual perception quality and image structure distortion, do not meet the requirements of clear and true medical image SR, and are not suitable for direct use in medical image SR. Therefore, this paper studies the SR method based on medical images.

To ensure that the medical images in the IoT-based intelligent healthcare system are clear and realistic, and can effectively assist doctors in diagnosing diseases. We propose a realistic medical image super-resolution (RMISR) method based on a pyramidal feature multiple distillation network. First, we constructed a novel pyramidal feature multi-distillation block (PFMDB). PFMDB gradually extracts features through cascading shallow residual blocks (SRBs), and uses pyramidal blocks (PYBs) to distill and extract refined features while expanding the receptive field. Second, we use PFMDB to build a lightweight SR network as a baseline model. Finally, on the basis of the baseline model, we increase the Laplacian gradient map branch to improve the visual perception quality of the SR image while ensuring the authenticity of the image structure. The innovations of this paper include the following: (1) A realistic medical image SR method based on pyramid feature multiple distillations is proposed, and a novel pyramid feature distillation block is designed, which can distill and extract refined features to improve medical image SR quality. (2) Based on the traditional digital image Laplacian operator, a two-branch SR network (SR branch and Laplacian gradient map branch) is designed, which can integrate the information of the Laplacian gradient map branch to ensure the real structure and improve the visual perception quality of the SR image. (3) We propose to use the contextual loss to optimize the high-frequency information of medical SR images in the Laplacian gradient map branch, use information entropy contrast-aware channel attention (ICCA) to assign different weights to the feature maps. (4) We propose to use two upsamplers in RMISR (sub-pixel upsampler and meta upsampler) and use an arbitrary scale upsampler to solve the problem of medical images requiring any multiple SR.

We discuss the related work of SR in Sect. 2. In Sect. 3, we define the SR process and baseline model of RMISR and describe in detail the medical image SR method

proposed in this paper in Sect. 4. Finally, we verify the effectiveness of RMISR through experiments in Sect. 5 and conclude in Sect. 6.

2 Related works

Image SR technology can recover the corresponding HR image from one or a series of related LR images [13, 14]. When the pathological part is very small (such as retinal images and coronary artery CT), or the resolution of medical images is low due to hardware equipment, shooting environment and other factors, medical image processing based on SR can reconstruct HR distillation help doctors clearly observe the lesion in intelligent healthcare systems [15]. The early SR methods were mostly based on interpolation, such as nearest-neighbor interpolation, bilinear interpolation, bicubic interpolation. The interpolation method uses the pixels around the sampling point to restore the target pixel. The bicubic interpolation method uses the gray values of 16 pixels around the sampling point for cubic interpolation and is often used to simulate the degradation process of HR images. Bicubic downsamples HR images, obtains corresponding LR images, and constructs pairs of images for training and testing of SR models based on deep learning. Before deep learning technology is applied to SR reconstruction, the most advanced SR methods are based on examples, such as sparse representation methods, which establish the mapping relationship between LR and SR images by learning the sparse association between image blocks [8]. The example-based SR method requires more preprocessing, which leads to low efficiency, and the nonlinear mapping from LR to high resolution is not optimized enough, which makes the effect in the non-uniform grayscale area of the SR image poor. Moreover, the reconstructed SR image has blurred visual perception and is not suitable for medical image SR.

The SR method based on a deep convolutional neural network can obtain a larger receptive field, learn more high-frequency detail information, and restore higher-quality SR images [16]. However, SRCNN did not achieve better results after increasing the number of network layers. The residual neural network can use the shortcut connection to directly connect the input LR image I^{LR} to the output HR image I^{SR} to achieve identity mapping. The SR model uses the residual network as the backbone network, which can increase the number of network layers while ensuring the convergence performance of the model. The SR methods super-resolution using very deep convolutional networks (VDSR) and EDSR proposed by Seoul National University in South Korea are both SR models

based on deep residual networks [17–19]. Liu Jie et al. proposed a residual feature aggregation network for image SR. They used a novel residual feature aggregation framework and enhanced spatial attention block and made full use of the layered features on the residual branches to improve the quality of SR reconstruction [20]. These methods have improved the objective evaluation indicators of SR (PSNR, SSIM) and achieved higher quality reconstruction results. However, VDSR, EDSR, RFASR and other pixel-based SR methods neglected visual perception when optimizing PSNR and SSIM, resulting in SR images that were too smooth. The SR method based on GAN can generate HR images with clear visual perception [21]. Ledig et al. [11] proposed a SR method based on generative adversarial network SRGAN, which used content loss and adversarial loss to improve the visual perception quality of generated pictures, and achieved good results. Self-Supervised photo upsampling via latent space exploration of generative models (PULSE) traverses the HR natural image manifold in a self-supervised manner, searches for images reduced to the original LR image. And it uses the characteristics of high-dimensional gaussian functions to limit the search space to ensure the authenticity of the output to generate SR images with clear visual perception [22]. Also, DBPN, a SR method based on the deep back-projection network, and Meta-SR, a SR method based on a dynamic prediction of the magnification filter weight, have achieved good results in high magnification SR and arbitrary magnification SR [23, 24]. The image reconstructed by the SR method based on pixel space optimization (VDSR, EDSR, FRASR) is too smooth, lacks detailed information, and is not clear enough. The image generated by the SR method based on feature space optimization (SRGAN, PULSE) has structural deformation and artifacts, which is not realistic enough. Therefore, these methods are not suitable for direct use in medical image SR reconstruction.

The IBM Research Institute of Australia has proposed a method of generating HR images from LR medical images using a progressive generation confrontation network. This method designs a multi-stage model that uses the triple loss function to use the output of the previous stage as a baseline to gradually improve the output image quality of the next stage. It can generate SR images with high scale factors while maintaining good image quality [25]. Xu Huang et al. proposed an SR reconstruction method for medical images based on GAN. The authors designed a medical pathology image preprocessing system to extract the image blocks of the tissue area, improve the discriminator by learning more prior information, and use the Huber loss function to replace the MSE loss function to improve the quality of the generated image [26]. Reference [27] proposed a medical image SR method based on dense

neural networks and a hybrid attention mechanism. This method removes the batch normalization layer in the dense neural network and adds a hybrid attention module to improve the high-frequency detail information of the SR image and improve the quality of the reconstructed image. Ma et al. [28] proposed a medical image SR method based on a relativistic average generative adversarial network, which helps to improve the quality of SR medical images in terms of numerical criteria and visual results. These SR methods have made improvements to the characteristics of medical images, but they lack a comprehensive consideration of the clarity and authenticity of the reconstructed image. For example, the medical image SR method based on the relativistic average generative adversarial network uses a feature space-based generative adversarial network to improve the clarity of visual perception. This method often causes structural distortions and even artifacts in the generated medical images, which is not conducive to the diagnosis of diseases by doctors.

To solve the above problems, this paper proposes a realistic medical image SR method based on a pyramidal feature multiple distillations network [29–31]. RMISR aims at reconstructing SR images with clear visual perception and real high-frequency detail information. By constructing a two-branch fusion SR network, the quality of visual perception is improved on the basis of ensuring the authenticity of the SR image structure [32, 33]. The SR branch learns the nonlinear mapping from LR images to HR images through PFMDB. Pyramidal feature multiple distillations can expand the receptive field while extracting refined features step by step, improving the reconstruction quality of SR images in pixel space [34–36]. The Laplacian gradient map branch learns the nonlinear mapping from the LR Laplacian gradient map to the HR Laplacian gradient map and fuses the SR Laplacian gradient map into the SR feature map to improve the visual perception quality of the SR image. We propose to use the traditional Laplacian gradient map to optimize the visual perception quality, which can avoid the structural deformation and distortion of the SR image caused by the feature space-based SR method, and make the SR image real and clear.

3 Preliminary overview

3.1 Process of image SR

The goal of image SR is to reconstruct HR images from LR images. Generally, the LR image I^{LR} is degenerated from the corresponding HR image, and the degradation process can be described by formula 1. Among them, I^{HR} is the corresponding HR image, $\mathcal{D}$ represents a degradation

mapping function, and θ represents the parameters of the degradation process such as noise and scale factor [37, 38]. To facilitate the acquisition of paired images for training, we use the bicubic ($\mathcal{B}$) to simulate the degradation process, as shown in formula 2, where ε represents the degradation scale factor. The process of SR can be described by formula 3, $\mathcal{S}$ represents the nonlinear mapping function from LR to HR, and δ represents the parameter of the function.

$$I^{LR} = \mathcal{D}(I^{HR}, \theta) \tag{1}$$

$$I^{LR} = \mathcal{B}(I^{HR}, \varepsilon), \tag{2}$$

$$I^{SR} = \mathcal{S}(I^{LR}, \delta), \tag{3}$$

The pixel space-based SR method aims to keep I^{SR} as consistent as possible with I^{HR} in pixels. The SR loss function based on pixel space is shown in formula 4, where $I_i^{HR} - I_i^{SR}$ calculates the pixel distance between HR image and SR image. Through continuous optimization of the network structure and training strategy, the performance on the objective evaluation indicators PSNR and SSIM is improving. However, PSNR tends to output results that are too smooth without enough details [12]. Although the pixel space-based SR method improves the PSNR index, the high-frequency information of the SR image is seriously lost and the visual perception is poor because the PSNR metric is fundamentally different from the subjective evaluation of human vision. Therefore, directly using the SR method based on pixel space to reconstruct medical images has difficulty meeting the requirements of medical images for rich detail information and clear texture information.

$$L_{MSE}^{SR} = \frac{1}{N}\sum_{i=1}^{N}|I_i^{HR} - I_i^{SR}|, \tag{4}$$

$$L_{VGG/i.j}^{SR} = \frac{1}{N}\sum_{x=1}^{N}|\phi_{i,j}(I_x^{HR}) - \phi_{i,j}(G(I_x^{LR}))|. \tag{5}$$

SR based on feature space is a perceptual driving method that aims to improve the visual perception quality of SR images. On the basis of GAN, it optimizes the feature space of SR models by using perceptual loss and counter-loss. The content loss function of the feature space is shown in formula 5, where $\phi_{i,j}()$ represents the visual geometry group (VGG) network that extracts features. Different from the pixel space-based SR method, the feature-based SR method uses the VGG network to extract the features of I^{HR} and I^{SR}, and aims to keep I^{SR} as consistent as possible with I^{HR} in the feature space, to encourage the generation of SR images with high visual perception. However, it is difficult to guarantee the authenticity of the SR image based on the feature space SR method, and the generated SR image often has artifacts or structural

deformation, which is not suitable for direct use in the SR restoration of medical images.

The application of deep learning technology in the medical field needs to fully measure the needs of society and people [39, 40]. The key requirement for SR reconstruction of medical images is clarity and reality. Based on the research of pixel space and feature space SR methods, we propose a medical image SR method combining traditional digital image processing and deep learning. In order to avoid the structural deformation caused by the feature space loss function, we use $L_{\mathrm{MSE}}^{\mathrm{SR}}$ in the SR branch to optimize the distance between I^{SR} and I^{HR} in the pixel space. To improve the image smoothing caused by the pixel space loss function, we combined the traditional image processing method to construct the Laplacian gradient map branch, and fused the Laplacian gradient image into the SR image. We build a two-branch network model to find a balance between pixel space and feature space-based SR methods, and reconstruct clear and true SR medical images.

3.2 RMISR baseline model

Medical images are an important basis for diagnosing diseases and are playing an increasingly important role in the medical field. Some medical images have low resolution due to small pathological parts or limitations of equipment technology and shooting environments. Doctors cannot clearly observe pathological parts through LR medical images, which is not conducive to disease diagnosis and treatment. To reconstruct clear and real medical images, this paper proposes a realistic medical image SR method based on a pyramidal feature multi-distillation network. We first designed a single-branch baseline model and then constructed a two-branch visual perception model based on the single-branch network. We have constructed a baseline model (RMISR-BL) to compare PSNR and SSIM with other methods to illustrate the advancement of RMISR-BL in pixel space SR reconstruction in this paper. Besides, RMISR-BL, as the basis of the two-branch SR network, lays the foundation for the reconstructed SR images with clear and realistic visual perception. The SR image reconstructed by RMISR has rich high-frequency information, clear texture details, and good visual perception quality. The baseline model of the RMISR method we propose is shown in Fig. 1:

The RMISR baseline model SR process includes three stages. First, we use a 3×3 convolutional layer (conv-3) to extract the initial features of medical images. Then, we use four PFMDB modules to construct a non-linear LR-to-SR mapping, use residual connections to reduce the computational complexity, and connect the output of PFMDB

of different depths to improve the detailed information of the feature map. Finally, we place a 1 * 1 convolution layer, two 3 * 3 convolution layers and a sub-pixel upsampler at the end of the network to reconstruct the SR image. Among them, we use 1 * 1 convolution to reduce the dimensionality and increase the nonlinearity to improve the SR reconstruction efficiency while ensuring the reconstruction quality. The sub-pixel upsampler is based on the depth information of feature maps, which can effectively improve the quality of upsampling. The PFMDB module includes SRB cascade components and PYB distillation components. We keep all the feature maps output by the cascade SRB components and copy all the feature maps for PYB multi-distillation to extract refined features (PFMDB details are in Sect. 4.2, pyramidal feature multi-distillation block).

4 Realistic medical image SR

4.1 Network architecture of RMISR

Our goal is to reconstruct clear and true HR images from LR medical images. We first constructed the RMISR baseline model (RMISR-BL). RMISR-BL is a lightweight model suitable for devices with low computing power and can be applied to primary hospitals lacking high-performance servers. To improve the quality of visual perception, we constructed a two-branch visual perception model (RMISR-VP). RMISR-VP is composed of SR branch and Laplacian gradient map branch, and its network structure is shown in Fig. 2.

The goal of the Laplacian gradient map branch is to learn the mapping from LR Laplacian gradient maps to HR maps. The Laplacian operator is an isotropic differential operator. We obtain the Laplacian gradient map of the image $I(x, y)$ by calculating the difference between the center pixel and the surrounding four pixels, as shown in formulas 6 and 7.

$$I_x(x, y) = I(x + 1, y) + I(x + 1, y) - 2I(x, y), \tag{6}$$

$$I_y(x, y) = I(x, y + 1) + I(x, y - 1) - 2I(x, y), \tag{7}$$

$$\nabla^2 I(x, y) = \left(I_x(x, y), I_y(x, y)\right). \tag{8}$$

Laplacian is a kind of differential operator. The Laplacian gradient map shows the gray-scale mutation area in the image $I(x, y)$, including high-frequency information such as edges and textures. The Laplacian gradient map contains information that is seriously missing in the SR reconstruction process. Therefore, fusing the information of the Laplacian gradient map branch can effectively improve the quality of the SR image. Define $\mathrm{Lap}(\cdot) = \nabla^2 I(x, y)$, which means the operation of

Fig. 1 RMISR baseline model network structure
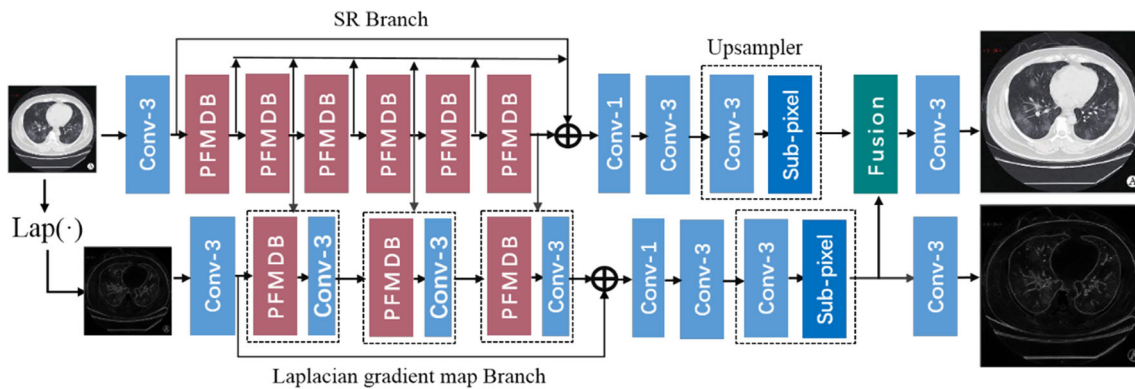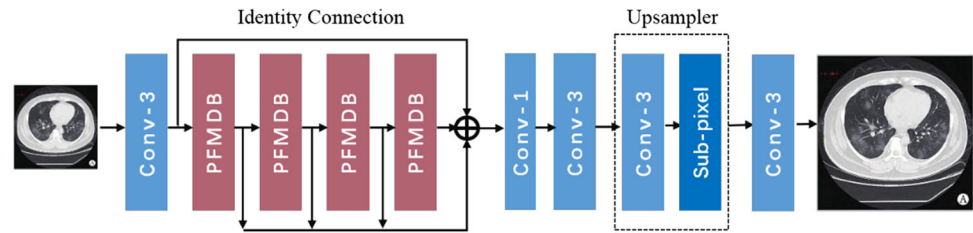
Fig. 2 RMISR visual perception model network structure

extracting the Laplacian gradient map. The Laplacian gradient map can be regarded as another kind of image, so SR can be used to learn the mapping between Laplacian gradient images from LR to SR mode. Most areas of the Laplacian gradient map are close to zero, and the convolutional neural network can focus on learning high-frequency information. Therefore, the Laplacian gradient map can more easily capture high-frequency information such as edges and textures, thereby improving the visual perception quality of SR images. As shown in Fig. 2, we fuse the feature representation of the middle layer of the SR branch in the Laplacian gradient map branch and use the SR branch feature map as prior knowledge to reduce the network depth and the number of parameters of the Laplacian gradient map branch. We fuse the Laplacian gradient feature map after the upsampling layer to the SR branch, optimize the high-frequency information of the SR map, and improve the edge and texture features of the image. Besides, we use a 3×3 convolutional layer to reconstruct the Laplacian gradient map.

SR branch is used to output the final SR image. To improve the quality of visual perception, the SR branch added two PFMDBs to the RMISR-BL. SR branch is divided into two parts. The first part is the same as the traditional SR method, mainly including 6 PFMDBs for learning the nonlinear mapping from LR to SR. The second part includes a fusion of SR features and Laplacian features and reconstruction of the final SR image. We enhance the edge and texture information by fusing the branch features of the Laplacian gradient map. The first part of the SR

branch uses 3×3 convolutional layers to extract 64 initial feature maps. Then, learn the detailed features of modal mapping from LR to SR by building 6 PFMDB modules. We connect the outputs of PFMDBs of different depths, merge the outputs of the 6 PFMDB modules, and merge the outputs of the second, fourth, and sixth PFMDB modules into the Laplacian filter branch. We use a sub-pixel upsampler to amplify the fusion features of the 6 PFMDB modules. The second part of the SR branch integrates the feature maps obtained from the Laplacian filter branch. We merge the feature maps of the two branches before feature map dimensionality reduction and use a 3×3 convolutional layer to reconstruct the final SR map. The RMISR method proposed in this paper uses the SR branch to learn the mapping of medical images from the LR mode to the SR mode in the pixel space to reconstruct the realistic and objective SR image with a high objective evaluation index. We use the Laplacian gradient map branch to learn the gradient map from LR mode to SR mode in pixel space and feature space to reconstruct the SR gradient map with rich texture information. Finally, we merge the information of the two branches to reconstruct a clear and undistorted medical SR image.

4.2 Pyramidal feature multi-distillation block

Inspired by the information multi-distillation block (IMDB) [29]. We designed the core component module PFMDB for the RMISR network. The internal structure of
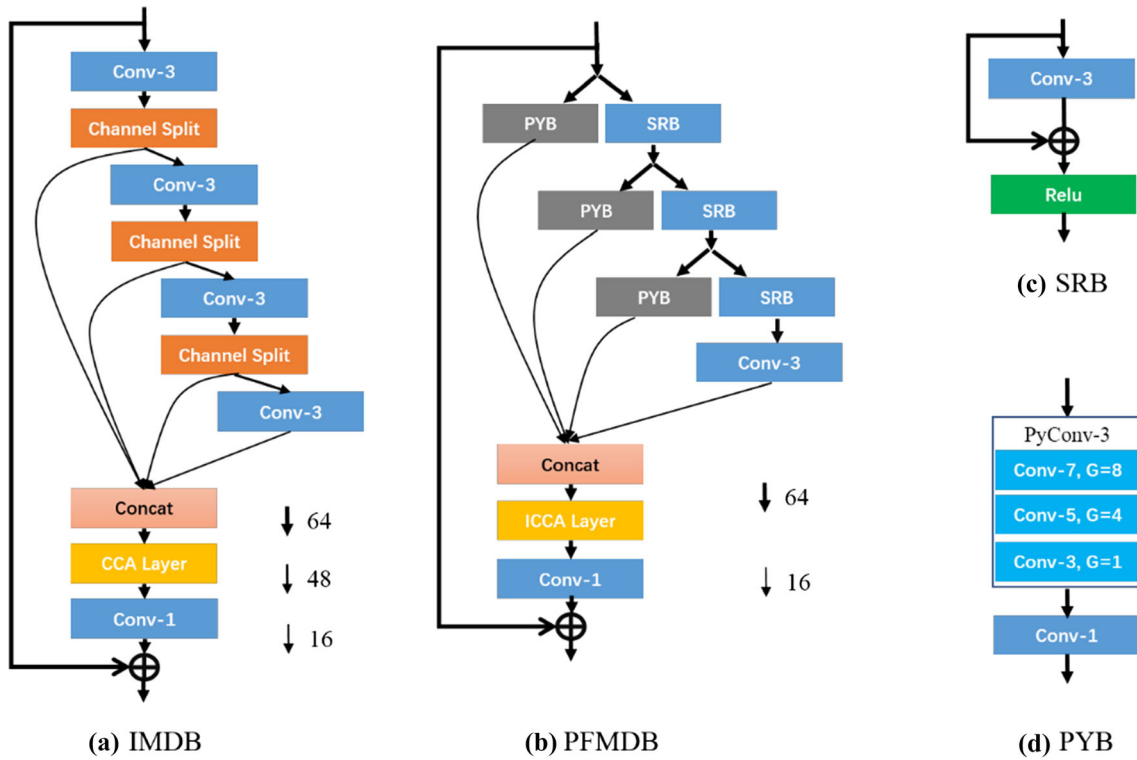
Fig. 3 The architecture of IMDB, PFMDB, SRB, and PYB: **a** information multi-distillation block. **b** Pyramidal feature distillation block, **c** shallow residual block and **d** pyramidal block

the module is shown in Fig. 3b. PFMDB includes two core components, SRB and PYB, as shown in Fig. 3c, d.

IMDB uses channel split to divide the 64 feature maps output by the convolutional layer into 2 groups by 1:3, of which 16 feature maps are directly input into the concat layer as distillation features, and the other 48 feature maps are used as the input of the next convolutional layer. We rethink IMDB's channel split strategy, by modifying the number of two grouping features, and adding a feature grouping selection algorithm to optimize the grouping strategy. Our progressive refinement module includes a pyramidal feature distillation module component and an SRB cascade component. The pyramidal feature distillation component is composed of three PYBs, and the SRB cascade component is composed of three SRBs and a Conv-3. Given the input feature F_{in}, the progressive refinement module processing flow can be described as:

$$DC_1(F_{dc1\_in}), CC_1(F_{cc1\_in}) = Split_1(L_1(F_{in})), \qquad (9)$$

$$DC_2(F_{dc2\_in}), CC_1(F_{cc2\_in}) = Split_2(L_2(F_{cc1_{out}})), \qquad (10)$$

$$DC_3(F_{dc3\_in}), CC_1(F_{cc3\_in}) = Split_3(L_3(F_{cc2_{out}})), \qquad (11)$$

$$DC_{4\_distilled} = L_4(F_{cc3\_out}), \qquad (12)$$

$$F_{distilled} = Concat(DC_1(F_{dc1_{in}}), DC_2(F_{dc2_{in}}), DC_3(F_{dc3_{in}}), DC_{4\_distilled}). \qquad (13)$$

As shown in formula 9–13, we first input the output features from the previous PFMDB to the cascade component (CC1) and make a copy into the distillation module (DC1). Second, we input the results of CC1 into the cascade component (CC2) and make a copy into the distillation module (DC2). Then, we input the results of CC2 to the cascade component (CC3) and make a copy into the distillation module (DC3). Finally, we concatenate the outputs of DC1, DC2, DC3, and CC3. The pyramidal block consists of a pyramid convolution (PyConv-3) and a 1×1 convolution. PyConv-3 contains three different scale convolution kernels 7×7, 5×5, and 3×3. This pyramid structure can directly expand the receptive field of RMISR, capture more context information, and extract different levels of detailed features [34]. Figure 3 shows that the input F_{in} of PyConv-3 is 64 feature maps. To expand the receptive field and reduce the number of model parameters and calculations, we introduce grouped convolution. Among them, 7×7 convolution inputs 64 feature maps and outputs 16 feature maps, divided into 8 groups of $G = 8$, 5×5 convolution inputs 64 feature maps, and outputs 16 feature maps, divided into 4 groups $G = 4$, 3×3 convolution inputs 64 feature maps and outputs 32 feature maps, without grouping $G = 1$. In the traditional 3×3 convolution, the input and output are 64 feature maps, and the number of parameters is 36,864. Under the

same conditions, the parameter amount of PyConv-3 is 31,104. Therefore, PYB can expand the receptive field while reducing the number of model parameters and calculations.

The SRB is composed of a 3×3 convolution, a residual connection, and a Relu activation function [30]. SRB adds a residual connection to the Conv-3 convolution of the cascade branch, which can use the residual connection to improve the learning ability from LR to SR mode, to make full use of residual learning without adding additional parameters. We use SRB cascade components to learn detailed features step by step, use PYB distillation components to extract distillation features, and measure the information and contrast of the feature map through the ICCA layer, and assign different weights to the feature map. SR is a low-level computer vision problem that focuses on the overall information of the image, such as brightness, contrast, and information entropy. Brightness is the average value of the gray value of the image, indicating the brightness of the image. The mean square error reflects the contrast of light and dark of the image. Information entropy is a measure that characterizes the amount of information. Entropy reflects the amount of information in an image, that is, the richness of details. We propose a channel attention mechanism ICCA based on information entropy (IE), standard deviation (SD), and mean (ME) [41]. Its structure is shown in Fig. 4 below (Fig. 5):

As shown in formula 14, we use IE, SD, and ME to replace global pooling, where the standard deviation and mean measure the contrast of the feature map, and information entropy measures the richness of the feature map. Define the input feature map of ICAA as $X = [x_1, x_2 \ldots x_c \ldots x_n]$, which means that there are $n = 64$ feature maps with space size $H \times W$, x_c represents the cth feature map, and the information entropy contrast channel attention is calculated as:
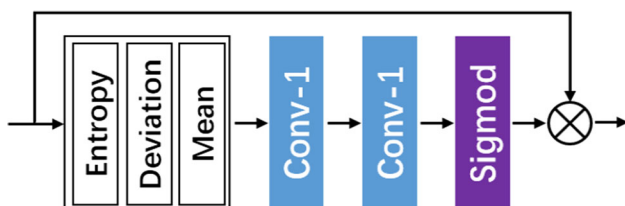

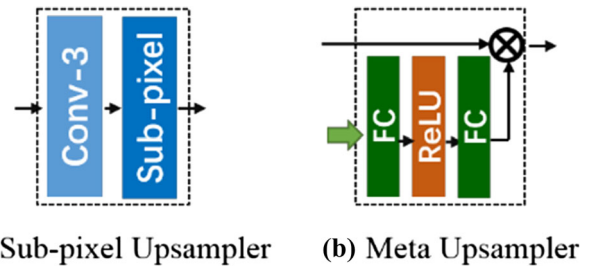**Fig. 4** Information entropy contrast-aware channel attention

(a) Sub-pixel Upsampler　　**(b) Meta Upsampler**

Fig. 5 The architecture of sub-pixel and meta upsampler

$$
s_c = \frac{1}{N} \sum_{(i,j) \in x_c} x_c^{i,j} \log_2 x_c^{i,j}
$$
$$
+ \sqrt{\frac{1}{HW} \sum_{(i,j) \in x_c} \left(x_c^{i,j} - \frac{1}{HW} \sum_{(i,j) \in x_c} x_c^{i,j} \right)^2}
$$
$$
+ \frac{1}{HW} \sum_{(i,j) \in x_c} x_c^{i,j}. \tag{14}
$$

where s_c represents the output of the cth feature map, and $N = 256$ is used to normalize the value of information entropy. ICCA assigns different weights to different channels by learning the contrast and information entropy of feature maps, which can effectively help RMISR to reconstruct SR images with clear visual perception.

4.3 Upsampling module of RMISR

The upsampling module in the SR model is a vital component. Different SR methods have different upsampling methods, and the positions of upsampling modules in the model are also different. SRCNN and VDSR use bicubic interpolation to first upsample the LR image, and then use the convolutional neural network to learn the nonlinear mapping from LR to HR to reconstruct the SR image. FSRCNN, EDSR, and RDN upsample the LR feature map at the end of the SR model. Medical images need to be upsampled at any scale. We use an upsampling unit based on position projection, weight prediction, and feature mapping [24].

We believe that each pixel of the medical SR image has the most relevant pixel on the LR image, and it is closely related to the magnification scale. To improve the quality of upsampling, we find the corresponding pixel (p'_x, p'_y) of each pixel (p_x, p_y) of the medical SR image in the LR image through position projection, where the position projection relationship can be expressed as formula 15. We reconstruct the medical SR image based on the most relevant pixels of LR. Then, we use a shallow fully connected network to predict the corresponding weight of each (p_x, p_y) of the medical SR image. The arbitrary scale upsampling module uses the shallow fully connected

network to dynamically predict the weight of the SR image based on the projection of the LR and SR pixel positions, according to the corresponding pixel offset and magnification. Based on the projection of the LR and SR pixel positions, we use the shallow fully connected network to dynamically predict the weight of the SR image according to the corresponding pixel offset and magnification scale, which can be expressed in formula 16.

$$(p_x', p_y') = \frac{p_x}{s}, \frac{p_y}{s}, \tag{15}$$

$$W(p_x, p_y) = f\left(\left(\frac{p_x}{s} - \frac{p_x}{s}, \frac{p_y}{s} - \frac{p_y}{s}\right); \theta\right), \tag{16}$$

$$I^{SR} = I^{LR}(p_x', p_y')W(p_x, p_y). \tag{17}$$

where f represents the weight prediction network, and θ is the weight of the weight prediction network. Based on the position projection and weight prediction, we use the matrix product to reconstruct the medical LR image into the corresponding SR image. The feature projection can be expressed in formula 17. The execution algorithm of the RMISR upsampling module can be described as follows:

The sub-pixel upsampler has a simpler structure than the arbitrary scale upsampler, with fewer parameters and high execution efficiency.

4.4 Loss functions of RMISR

We propose an optimization method using L1 loss and contextual loss. Different loss function combination optimization schemes are designed and the effectiveness of the proposed method is verified through experiments. The loss function proposed in this paper is shown in formula 18, where L_T represents total loss, L_{MAE} represents mean absolute error (MAE) loss, L_C represents contextual loss, and α and β are coefficients used to balance different losses. When calculating the loss function, we obtain the LR image I^{LR} by downsampling the given HR image I^{HR} using bicubic interpolation. The SR image is obtained from the RMISR network $I^{SR} = F_{RMISR}(I_i^{LR})$, where $F_{RMISR}(\cdot)$ represents the RMISR network we built. Given the training set $\{I_i^{LR}, I_i^{HR}\}_{i=1}^N$, the MAE loss function uses the L1 paradigm to compare the difference between I^{HR} and I^{SR}

Algorithm 1 RMISR arbitrary scale upsample module algorithm

1: **Input:** I^{LR}, s # input LR image and scale factor

2: **Output:** I^{SR} # output SR image

3: Define (p_x, p_y) as I^{SR} pixels

4: Define H, W as I^{SR} height and width

5: **for** x=0: H do

6:      **for** y=0: W do

7:          $(p_x', p_y') = \left\lfloor \frac{p_x}{s}, \frac{p_y}{s} \right\rfloor$ #The most relevant pixels of LR and SR

8:          $W(p_x, p_y) = f\left(\left(\frac{p_x}{s} - \left\lfloor\frac{p_x}{s}\right\rfloor, \frac{p_y}{s} - \left\lfloor\frac{p_y}{s}\right\rfloor\right); \theta\right)$ #Weight of the most relevant pixel mapping

9:          $I^{SR} = I^{LR}(p_x', p_y')W(p_x, p_y)$

10:      **end for**

11: **end for**

12: return I^{SR}

The time complexity of the whole algorithm is $O(H * W)$, where H is the height of the SR image, W is the width of the SR image. In addition to the arbitrary scale upsampler, we also use sub-pixel upsamplers in RMISR-BL and RMISR-VP. The sub-pixel up-sampler includes a convolutional layer and a sub-pixel convolutional layer, which is different from adding a large number of zeros during the up-sampling of the deconvolution layer. The sub-pixel layer regards the pixels on the feature map as sub-pixels of the SR image. All the pixels on the feature map are combined to reconstruct the pixels on the SR map.

pixel by pixel, which can be expressed by formula 19.

$$L_T = \alpha L_{MAE} + \beta L_C, \tag{18}$$

$$L_{MAE} = \frac{1}{N} \sum_{i=1}^N \|F_{RMISR}(I_i^{LR}) - I_i^{HR}\|_1, \tag{19}$$

$$L_C(\phi(I^{SR}), \phi(I^{HR}), l) = -\log(CX(\phi^l(I^{SR}), \phi^l(I^{HR}))). \tag{20}$$

Perceptual loss is used to measure the perceptual similarity of two images, by mapping SR and HR images to the same feature space and calculating the distance between

feature maps [42]. The perceptual loss function measures the distance between the I^{SR} and I^{HR} feature maps, and is used to constrain the similarity of global high-frequency features. The lack of constraints on the similarity of local features may lead to structural deformation of the I^{SR}. Therefore, from the perspective of a probability distribution, we hope that the feature distribution between the SR image and the target HR image is as similar as possible. The context loss L_C can measure the similarity between the features of the SR image I^{SR} and the target image I^{HR}. By comparing the similarity of local features, it can ensure that the visual perception of the restored SR image is clear and natural. The context loss used in this paper is shown in formula 20, where CX_{ij} is the similarity of local features x_i and y_j. By optimizing the distance measurement between the local feature distribution of I^{SR} and I^{HR}, it is ensured that the restored SR image is closer to the real HR image in the feature space, and the goal of a clear and true medical SR image is achieved. Our SR branch uses the L1 loss function to optimize the SR image in the pixel space to ensure the authenticity of the image. Laplacian gradient map branch uses L1 loss function and contextual loss, where the coefficients of L1 loss and contextual loss are 10 and 1, respectively. Laplacian gradient map branch can reconstruct a gradient map with rich high-frequency information. We can reconstruct the realistic SR medical image with clear visual perception by fusing the feature maps of the two branches.

5 Experimental results

5.1 Implementation details

In the experiment, we use the DIV2K dataset as the training set and Set5, Set14, BSD100, Urban100, and Manga109 as the test set. In the same software and hardware environment, using the same data set and setting the same training rounds, we trained the two network models we designed RMISR-BL, RMISR-VP, and some comparison method network models, such as SRCNN, FSRCNN, VDSR, EDSR. We randomly crop the HR images in the DIV2K training set to 192×192 as the input of the SR model and set the mini-batch size to 16. Two RMISR models use the ADAM optimizer and set the momentum parameter $\beta_1 = 0.9, \beta_2 = 0.999$. The initial learning rate is set to 2×10^{-4}, and halved at every 2×10^2 iterations. We use a pre-trained VGG-19 model in the contextual loss function to measure the similarity of features between SR images and HR images. We use PyTorch to build a realistic medical image SR model. In the model training phase, we use a high-performance server to train and verify the

effectiveness of our model. The high-performance server with Linux operating system and the GPU is NVIDIA Tesla T4.

SR methods based on pixel space generally use PSNR and SSIM as evaluation indicators. The higher the values of PSNR and SSIM are, the better the quality of the SR image. PSNR and SSIM are relatively sensitive to image brightness, structure, contrast, pixel location, and are objective evaluation indicators widely used in the SR field. However, the PSNR and SSIM evaluation methods do not match the human visual perception to a high degree. Therefore, this paper uses PSNR and SSIM to evaluate the pixel space-based SR method RMISR-BL, and mean opinion score (MOS) testing to evaluate the visual perception-based SR method RMISR-VP.

5.2 Experimental results of RMISR-BL

Through experimental observations, the model that uses the L1 or L2 loss function as the optimization target in the SR method will cause the high-frequency information of the reconstructed medical SR image to be too smooth, and the overall visual perception is blurred. However, based on the pixel space SR method, the higher the PSNR value, the better the visual perception quality. Therefore, we first constructed the SR model RMISR-BL based on the L1 loss function. The goal is to construct a model with a small number of parameters and a high PSNR value. Then, based on RMISR-BL, the Laplacian filter branch is used to enrich the high-frequency information of medical SR images, improve the details and texture characteristics, and improve the quality of visual perception. We selected 15 comparison methods, including bicubic interpolation and MemNet, ESPCN, CARN-M and so on. The comparison results are shown in Table 1 [43–45].

It can be seen from Table 1 that the RMISR-BL proposed in this paper has achieved the best performance on the BSDS and Urban100 datasets, and PSNR and SSIM are higher than the other 15 comparison methods. On the Set5 and Set14 data sets, the PSNR and SSIM values of RMISR-BL are very close to SelNet, but the parameter of RMISR-BL is 0.848 M, which is much smaller than SelNet's 1.417 M. On the Manga109 dataset, the PSNR and SSIM values of RMISR-BL rank second, second only to the comparison method SRMDNF. The parameter amount of RMISR-BL is approximately half of SRMDNF. It should be noted that our RMISR-BL in Table 1 is only trained on the DIV2K dataset for 5000 rounds, which is far less than the training times of other comparison methods.

We selected 1 to 1000 rounds of training data, including the L1 loss value and PSNR value. A total of 109 data points, including [1, 2, 3… 100] and [200, 300… 1000] are used in the chart. Figure 6 shows that the SRCNN

Table 1 Performance comparison with state-of-the-art algorithms for × 4 image super-resolution

SR algorithms	Params (M)	Scale	Set5		Set14		BSDS100		Urban100		Manga109	
			PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	–	4	28.42	0.810	26.10	0.702	25.96	0.667	23.15	0.657	24.92	0.789
SRCNN [9]	0.057	4	30.48	0.863	27.49	0.750	26.90	0.710	24.52	0.722	27.66	0.851
FSRCNN [10]	0.012	4	30.71	0.866	27.59	0.754	26.98	0.715	24.62	0.728	27.90	0.852
VDSR [13]	0.665	4	31.35	0.884	28.01	0.767	27.29	0.725	25.18	0.752	28.83	0.881
DRCN [19]	1.774	4	31.53	0.885	28.02	0.767	27.23	0.723	25.14	0.751	28.98	0.882
LapSRN [35]	0.502	4	31.54	0.885	28.09	0.770	27.32	0.726	25.21	0.756	29.09	0.890
CNF [33]	0.337	4	31.55	0.8856	28.15	0.768	27.32	0.7253	–	–	–	–
DRRN [13]	0.297	4	31.68	0.889	28.21	0.772	27.38	0.728	25.44	0.764	29.46	0.896
BTSRN [14]	0.410	4	31.85	–	28.20	–	27.47	–	25.74	–	–	–
MemNet [43]	0.667	4	31.74	0.889	28.26	0.772	27.40	0.728	25.50	0.763	–	–
ESPCN [44]	–	4	29.21	0.851	26.40	0.744	25.50	0.696	24.02	0.726	23.55	0.795
SRGAN [11]	1.6	4	29.46	0.838	26.60	0.718	25.74	0.666	24.50	0.736	27.79	0.856
SRMDNF [37]	1.555	4	31.96	0.893	28.35	0.777	**27.49**	**0.734**	**25.68**	**0.773**	*30.12*	*0.902*
SelNet [38]	1.417	4	*32.00*	0.893	*28.49*	**0.778**	27.44	0.733	–	–	–	–
CARN-M [45]	0.412	4	31.92	0.890	28.42	0.776	27.44	0.730	25.62	0.769	–	–
RMISR-BL	0.848	4	**31.96**	**0.890**	**28.44**	*0.780*	27.49	*0.735*	25.77	*0.775*	**30.08**	**0.900**

The italic number indicates the best result and the bold number indicates the second best result. "–" denotes the results that are not reported

converged stably to 0.0054 within 100 rounds of training, and the PSNR value rose to 25.57 db. The loss value of the baseline model RMISR-BL and information multi-distillation network (IMDN) did not fully converge in 1000 rounds of training, and the PSNR value gradually increased during the training process. The PSNR value of RMISR-BL was the highest at 1000 rounds, reaching 30.11 db. Because the PSNR value of SRCNN is much lower than IMDN and RMISR-BL, to facilitate the display, we replaced 109 data points with the highest PSNR value of SRCNN. Besides, by introducing additional data sets (Flickr2K, OST, etc.) and increasing the number of training rounds, we can further improve the PSNR and SSIM values of RMISR-BL. It can be found from Table 1 and Fig. 5 that our RMISR-BL obtains a better pixel space

reconstruction effect than the state-of-the-art SR method. Based on RMISR-BL, the construction of a double-branch structure SR network can effectively ensure the authenticity of SR images and play an important role in improving the clarity of visual perception of SR images.

5.3 Experimental results of RMISR-VP

We designed the SR branch of RMISR-VP on the basis of RMISR-BL, and improved the quality of visual perception by fusing information from the Laplacian gradient map branch. Medical image SR reconstruction improves visual perception while ensuring the authenticity of medical SR images. Therefore, the SR branch in this paper includes 6 PFMDBs, and 2 more PFMDBs than RMISR-BL, and the

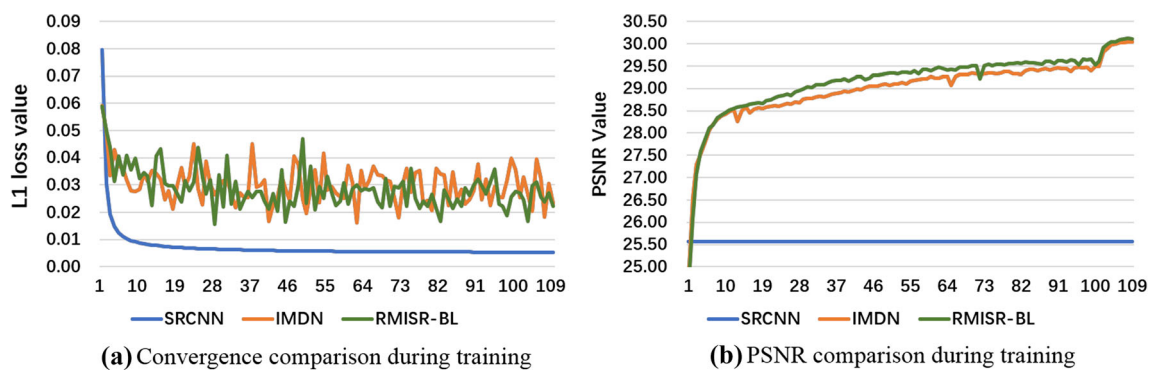**(a)** Convergence comparison during training **(b)** PSNR comparison during training

Fig. 6 Convergence and PSNR comparison of different SR methods during training

rest of the network structure is the same as RMISR-BL. By adding two PFMDBs, the PSNR value of the SR branch is increased by 0.1 ± 0.02 db compared with RMISR-BL, which further improves the quality of SR image reconstruction based on L1 loss function optimization. The Laplacian gradient map branch contains 3 PFMDBs, and the output feature maps of the second, fourth, and sixth PFMDB of the SR branch are merged to reduce the number of parameters and improve the quality of reconstruction. To improve the quality of high-frequency detail features such as edges and textures, we have added a contextual loss function to the L1 loss function. The CT Laplacian gradient map of the lung reconstructed by different SR methods is shown in Fig. 7. We choose SRCNN, SRGAN, and IMDN as the comparison methods of visual effects. SRCNN and IMDN are SR methods based on pixel space, and SRGAN is an SR method based on feature space. We choose these methods to compare our approach from the two dimensions of pixel space and feature space and verify clear and realistic SR reconstruction targets. We adopt the commonly used scale factor ($\times$ 4) by both the Laplacian gradient map branch and the SR branch to reconstruct the lung CT image.

RMISR-L1 represents the Laplacian gradient map branch that only uses L1 as the loss function. RMISR-CX represents the Laplacian gradient map branch that only uses contextual loss as the loss function, and RMISR-L1 + CX represents the Laplacian gradient map branch that uses L1 and contextual loss as the loss function. It can be seen from Fig. 7 that when the loss function is L1, the medical SR image reconstructed by our Laplacian gradient map branch is clearer than other methods (Bicubic, SRCNN, VDSR, IMDN), and the edge texture is sharper. However, like other SR methods based on the L1 loss function, the tiny edges and textures are smoothed out during the reconstruction process and almost all are lost. The medical SR image reconstructed by the Laplacian gradient map branch based on the contextual loss function has a wealth of high-frequency information, including a large amount of tiny edge and texture information. To ensure the authenticity of high-frequency information, we comprehensively apply L1 loss and contextual loss, set $\alpha = 10$ and $\beta = 1$ in formula 16, to reconstruct a more natural medical image.

It can be seen from Fig. 8 that by fusing the information of the Laplacian gradient map branch, the lung CT image reconstructed by RMISR-VP has rich high-frequency information such as edges and textures, and the visual perception is natural and clear. RMISR-VP is based on the traditional Laplacian operator, and the reconstruction of high-frequency information does not rely on GAN. Therefore, the medical SR image reconstructed by RMISR-VP does not have structural deformation and artifacts. To

verify the visual perception effect of RMISR-VP reconstructed medical images, we cooperated with dermatologists at Xiangya Hospital to reconstruct 100 demodicosis, 100 flat warts, and 100 seborrheic keratosis dermoscopy images. The SR reconstruction methods include Bicubic, SRCNN, SRGAN, IMDN, RMISR-BL, and RMISR-VP. We set the 300 SR medical images reconstructed by Bicubic to 1 (bad quality) and 300 HR images to 5 (excellent quality). Sixteen raters were asked to rate the SR images from 1 to 5. According to statistics, the average score of SRCNN is 2.2, SRGAN is 3.7, IMDN is 3.3, RMISR-BL is 3.4, and RMISR-VP is 4.1. Mean opinion score testing further proves the effectiveness of the proposed SR method.

5.4 Experimental results of RMISR arbitrary magnification

Traditional SR methods generally include $\times$ 2, $\times$ 3, and $\times$ 4 scale SR reconstruction, which lack flexibility. Medical image SR reconstruction needs to be reconstructed at any scale according to the actual application. As shown in Fig. 5, this paper uses a sub-pixel upsampler and an arbitrary scale upsampler. According to the specific situation of the hospital, our SR model can choose different upsamplers. To improve computational efficiency and reduce the amount of parameters, a sub-pixel upsampler can be used. To solve the problem of SR reconstruction of medical images at any scale, we replaced the sub-pixel upsampler in RMISR-VP with an arbitrary scale upsampler, and trained an RMISR-VP model that can realize SR reconstruction at any scale. The arbitrary scale upsampler is an up-sampling unit based on position projection, weight prediction and feature mapping. By learning the most relevant pixels of LR and SR, the medical SR image can be reconstructed at any multiple.

The results of SR reconstruction of retinopathy images at any scale are shown in Fig. 9. RMISR-VP can reconstruct SR images of any scale according to the needs of doctors. We use fractional scale factors ($\times$ 2.5, $\times$ 3.5, $\times$ 4.5) that are not commonly used in other SR methods to reconstruct retinopathy images and also display integer scale factor ($\times$ 5) reconstructed SR image. The medical images reconstructed by RMISR-VP have real and clear visual perception. We can see from the zoomed-in details that the retinopathy texture reconstructed by the fractional scale factor and the integer scale factor is not only visually clear but also free of structural deformation and artifacts. This flexible SR reconstruction strategy can effectively assist doctors in diagnosing the disease. To facilitate doctors' use of the RMISR model to reconstruct medical images, and promote the promotion of deep learning models in the medical field. We use PyQt5 to
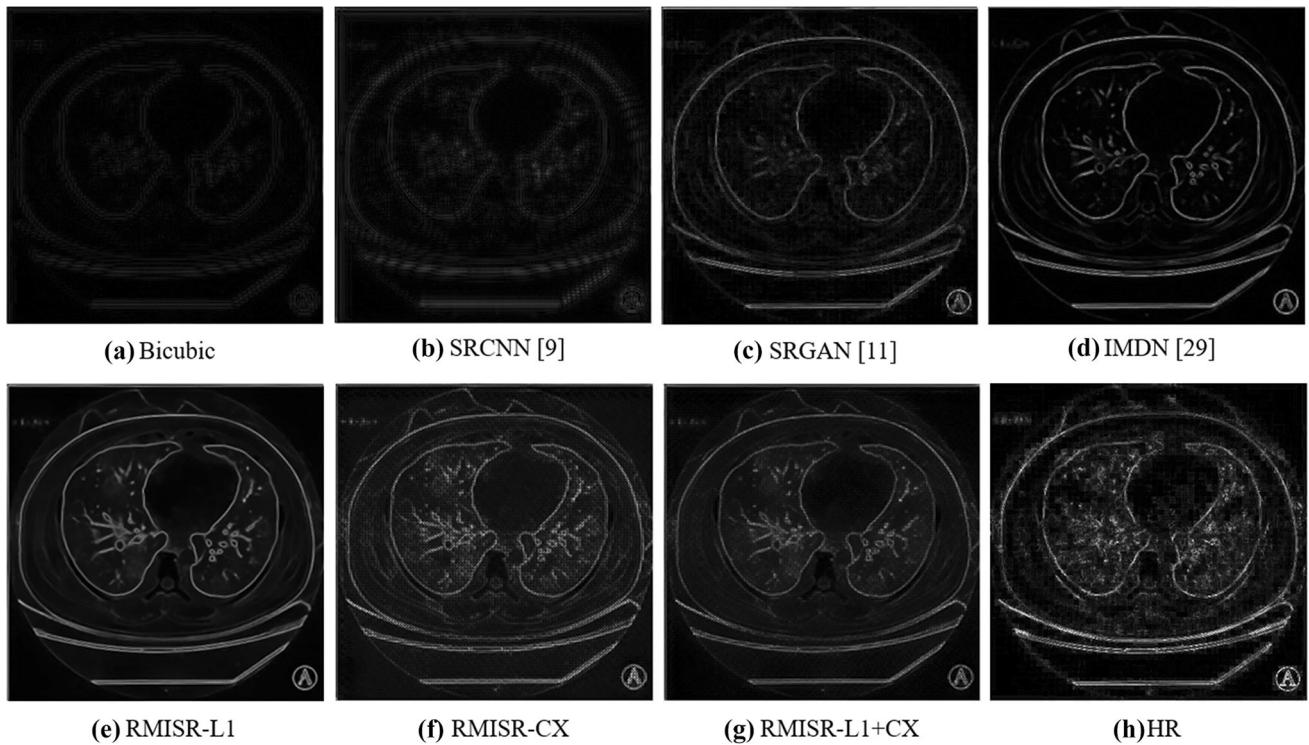
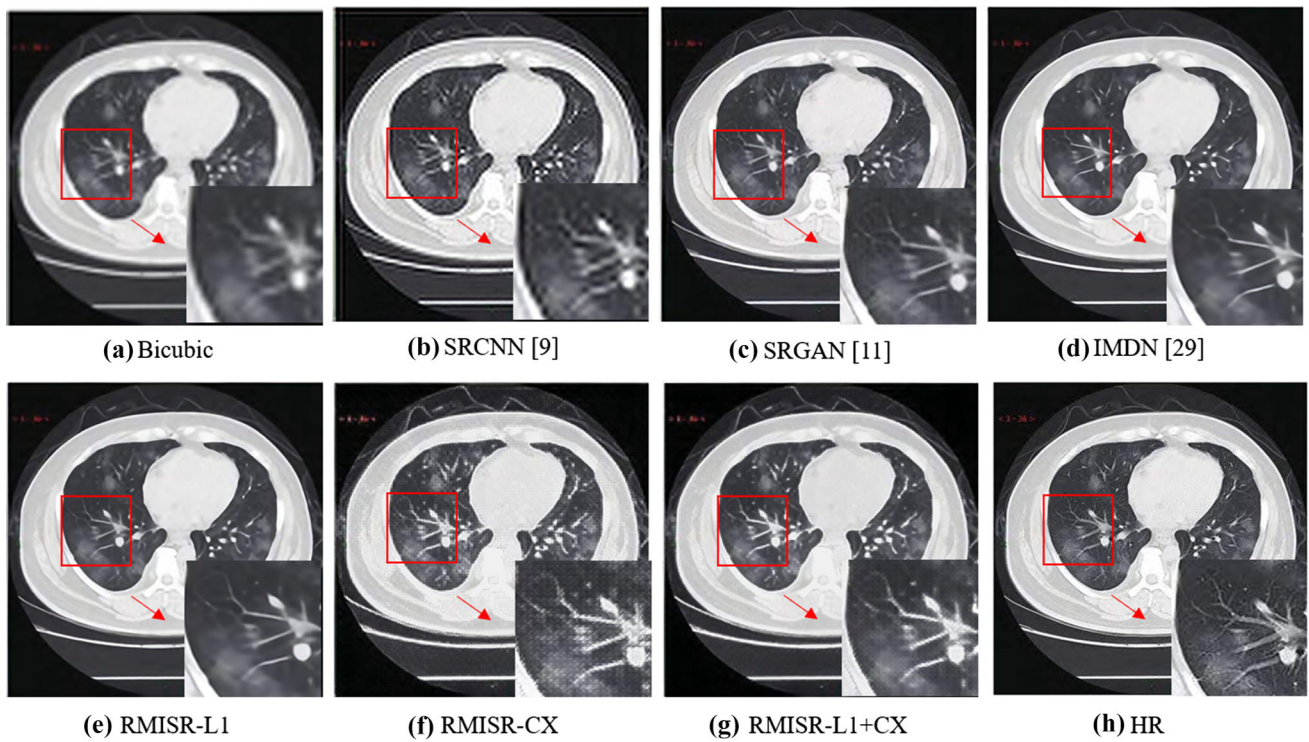**Fig. 7** Comparison of Laplacian gradient maps reconstructed by different SR methods

Fig. 8 Comparison of lung CT images reconstructed by different SR methods

encapsulate the trained RMISR-BL, RMISR-VP and other SR models. By designing a simple and easy-to-use interface, doctors can use the deep learning-based SR model to perform SR reconstruction on medical images with a simple mouse operation. To improve the efficiency of SR reconstruction, we designed the image area selection
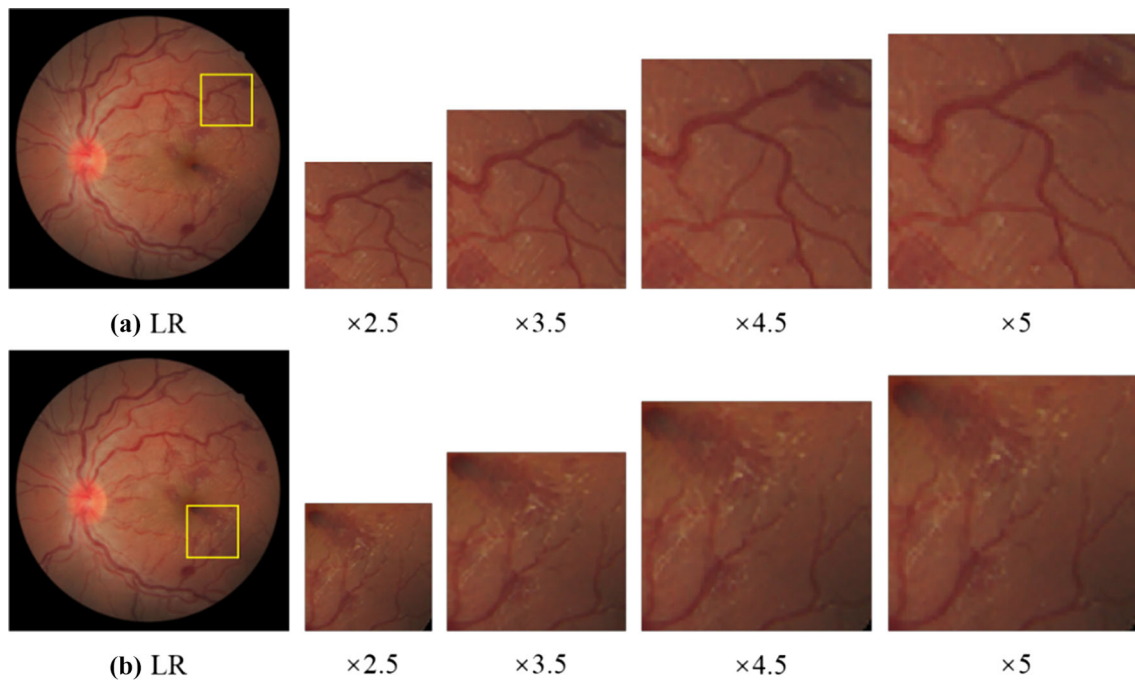
Fig. 9 SR reconstruction of retinopathy images at any scale

function according to the doctor's needs. The doctor can only perform SR reconstruction on the key areas of the LR image (such as the lesion) through the operation of the mouse. Besides, we also designed SR reconstruction multiple selection, SR algorithm selection and SR image quality analysis functions.

6 Conclusion

To solve the problem of real and clear SR reconstruction of medical images in IoT-based intelligent healthcare systems, this paper proposes the RMISR method. First, we designed a core module PFMDB based on PFB and SRB. PFMDB gradually extracts feature maps by cascading SRB, and uses PFB distillation to extract different levels of detailed features. By calculating the information entropy and contrast of the feature map, we designed a channel attention mechanism ICCA, which assigns different weights to the output feature map of PFMDB. Then, we constructed the baseline model RMISR-BL based on PFMDB. Comparative experiments show that RMISR-BL is superior to other methods in PSNR and SSIM. Finally, on the basis of RMISR-BL, we constructed a two-branch visual perception model RMISR-VP, which improves the edge and texture characteristics of medical SR images by fusing the information of the Laplacian gradient map branch, and improves the quality of visual perception. Besides, we propose to use the L1 loss function and contextual loss function in the Laplacian gradient map branch.

Then we use an arbitrary scale upsampler to achieve any scale SR reconstruction of medical images and develop a medical auxiliary diagnosis component for IoT-based intelligent healthcare systems, which is convenient for doctors to use deep learning-based models to perform medical image SR reconstruction. RMISR can effectively solve the adverse effects of LR images on IoT-based intelligent healthcare systems and assist doctors in diagnosing diseases. Our next step is to study the lightweight medical image SR method.

Availability of data and materials Not applicable.

Code availability Not applicable.

Declaration

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Guo K, Ren S, Bhuiyan MZA et al (2019) MDMaaS: medical-assisted diagnosis model as a service with artificial intelligence and trust. IEEE Trans Ind Inf 16(3):2102–2114
2. Wang X, Yang LT, Wang Y, Ren L, Deen MJ (2020) ADTT: a highly-efficient distributed tensor-train decomposition method for IIoT big data. IEEE Trans Ind Inf. https://doi.org/10.1109/TII.2020.2967768
3. Zhou X, Liang W, Kevin I et al (2020) Deep-learning-enhanced human activity recognition for Internet of healthcare things. IEEE Internet Things J 7(7):6429–6438
4. Hao F, Pei Z, Yang LT (2020) Diversified top-k maximal clique detection in social internet of things. Future Gener Comput Syst 107:408–417
5. Long E, Lin H, Liu Z et al (2017) An artificial intelligence platform for the multihospital collaborative management of congenital cataracts. Nat Biomed Eng 1(2):1–8
6. Wang X, Yang LT, Song L, Wang H, Ren L, Deen J (2020) A tensor-based multi-attributes visual feature recognition method for industrial intelligence. IEEE Trans Ind Inf. https://doi.org/10.1109/TII.2020.2999901
7. Ren S, Jain DK, Guo K, Xu T, Chi T (2019) Towards efficient medical lesion image super-resolution based on deep residual networks. Signal Process Image Commun 75:1–10
8. Yang J, Wright J, Huang TS et al (2010) Image super-resolution via sparse representation. IEEE Trans Image Process 19(11):2861–2873
9. Dong C, Loy CC, He K et al (2014) Learning a deep convolutional network for image super-resolution. In: European conference on computer vision. Springer, Cham, pp 184–199
10. Dong C, Loy CC, Tang X (2016) Accelerating the super-resolution convolutional neural network. In: European conference on computer vision. Springer, Cham, pp 391–407
11. Ledig C, Theis L, Huszár F et al (2017) Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4681–4690
12. Wang X, Yu K, Wu S et al (2018) Esrgan: enhanced super-resolution generative adversarial networks. In: Proceedings of the European conference on computer vision, pp 1–17
13. Tai Y, Yang J, Liu X (2017) Image super-resolution via deep recursive residual network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3147–3155
14. Fan Y, Shi H, Yu J et al (2017) Balanced two-stage residual networks for image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 161–168
15. Zhou X, Li Y, Liang W (2020) CNN-RNN based intelligent recommendation for online medical pre-diagnosis support. IEEE/ACM Trans Comput Biol Bioinf. https://doi.org/10.1109/TCBB.2020.2994780
16. Shamsolmoali P, Zareapoor M, Jain DK, Jain VK, Yang J (2019) Deep convolution network for surveillance records super-resolution. Multimed Tools Appl 78(17):23815–23829
17. Kim J, Kwon LJ, Mu LK (2016) Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1646–1654
18. Lim B, Son S, Kim H et al (2017) Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 136–144
19. Kim J, Kwon LJ, Mu LK (2016) Deeply-recursive convolutional network for image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1637–1645
20. Liu J, Zhang W, Tang Y, Tang J, Wu G (2020) Residual feature aggregation network for image super-resolution. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 2359–2368
21. Jain DK, Zareapoor M, Jain R, Kathuria A, Bachhety S (2020) GAN-Poser: an improvised bidirectional GAN model for human motion prediction. Neural Comput Appl 32(18):14579–14591
22. Menon S, Damian A, Hu S, Ravi N, Rudin C (2020) PULSE: self-supervised photo upsampling via latent space exploration of generative models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 2437–2445
23. Haris M, Shakhnarovich G, Ukita N (2018) Deep back-projection networks for super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1664–1673
24. Hu X, Mu H, Zhang X et al (2019) Meta-SR: a magnification-arbitrary network for super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1575–1584
25. Mahapatra D, Bozorgtabar B, Garnavi R (2019) Image super-resolution using progressive generative adversarial networks for medical image analysis. Comput Med Imaging Graph 71:30–39
26. Huang X, Zhang Q, Wang G et al (2019) Medical image super-resolution based on the generative adversarial network. In: Chinese intelligent systems conference. Springer, Singapore, pp 243–253
27. Liu K, Ma Y, Xiong H et al (2019) Medical image super-resolution method based on dense blended attention network. arXiv:1905.05084
28. Ma Y, Liu K, Xiong H et al (2021) Medical image super-resolution using a relativistic average generative adversarial network. Nucl Instrum Methods Phys Res Sect A 992(165053):1–6
29. Hui Z, Gao X, Yang Y et al (2019) Lightweight image super-resolution with information multi-distillation network. In: Proceedings of the 27th acm international conference on multimedia, pp 2024–2032
30. Liu J, Tang J, Wu G (2020) Residual feature distillation network for lightweight image super-resolution. arXiv:2009.11551
31. Zhou X, Liang W, Shimizu S, Ma J, Jin Q (2020) Siamese neural network based few-shot learning for anomaly detection in industrial cyber-physical systems. IEEE Trans Ind Inf 17(8):5790–5798
32. Ma C, Rao Y, Cheng Y et al (2020) Structure-preserving super resolution with gradient guidance. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 7769–7778
33. Ren H, El-Khamy MLJ (2017) Image super resolution based on fusing multiple convolution neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 54–61
34. Duta IC, Liu L, Zhu F et al (2020) Pyramidal convolution: rethinking convolutional neural networks for visual recognition. arXiv:2006.11538
35. Lai WS, Huang JB, Ahuja N et al (2017) Deep Laplacian pyramid networks for fast and accurate super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 624–632

36. Zareapoor M, Jain DK, Yang J (2018) Local spatial information for image super-resolution. Cogn Syst Res 52:49–57

37. Zhang K, Zuo W, Zhang L (2017) Learning a single convolutional super-resolution network for multiple degradations. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3262–3271

38. Choi JS, Kim M (2017) A deep convolutional neural network with selection units for super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 154–160

39. Hao F, Pang G, Wu Y et al (2019) Providing appropriate social support to prevention of depression for highly anxious sufferers. IEEE Trans Comput Soc Syst 6(5):879–887

40. Zhou X, Xu X, Liang W et al (2021) Intelligent small object detection based on digital twinning for smart manufacturing in industrial CPS. IEEE Trans Ind Inf. https://doi.org/10.1109/TII.2021.3061419

41. Jain DK, Kumar A, Garg G (2020) Sarcasm detection in mash-up language using soft-attention based bi-directional LSTM and feature-rich CNN. Appl Soft Comput 91:106198

42. Mechrez R, Talmi I, Shama F et al (2018) Maintaining natural image statistics with the contextual loss. In: Asian conference on computer vision. Springer, Cham, pp 427–443

43. Tai Y, Yang J, Liu X et al (2017) Memnet: a persistent memory network for image restoration. In: Proceedings of the IEEE international conference on computer vision, pp 4539–4547

44. Shi W, Caballero J, Huszár F et al (2016) Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1874–1883

45. Timofte R, Gu S, Wu J et al (2018) Ntire 2018 challenge on single image super-resolution: methods and results. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 852–863